UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 10/607,667 | 06/27/2003 | Boris Chidlovskii | D/A2470 | 8045 |

25453 7590 03/02/2006

PATENT DOCUMENTATION CENTER
XEROX CORPORATION
100 CLINTON AVE., SOUTH, XEROX SQUARE, 20TH FLOOR
ROCHESTER, NY 14644

| EXAMINER | |
|---|---|
| DEBROW, JAMES J | |
| **ART UNIT** | **PAPER NUMBER** |
| 2176 | |

DATE MAILED: 03/02/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

| Office Action Summary | Application No. | Applicant(s) | |
|---|---|---|---|
| | 10/607,667 | CHIDLOVSKII ET AL. | |
| | Examiner | Art Unit | |
| | James J. Debrow | 2176 | |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on *27 June 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) *1-20* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-20* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *6/27/2003* is/are: a)☐ accepted or b)☒ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date *6/27/2003*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

## DETAILED ACTION

1. This action is responsive to communications: Application filed on 27 Jun 2003

2. Claims 1-20 are pending in the case. Claims 1, and 11 are independent claims.

### *Drawings*

3. The drawings are objected to as failing to comply with 37 CFR 1.84(p)(5) because they do not include the following reference sign(s) mentioned in the description: **item 20, 22, and 26 on pages 15 and 16, in reference to Figure 5**. Corrected drawing sheets in compliance with 37 CFR 1.121(d) are required in reply to the Office action to avoid abandonment of the application. Any amended replacement drawing sheet should include all of the figures appearing on the immediate prior version of the sheet, even if only one figure is being amended. Each drawing sheet submitted after the filing date of an application must be labeled in the top margin as either "Replacement Sheet" or "New Sheet" pursuant to 37 CFR 1.121(d). If the changes are not accepted by the examiner, the applicant will be notified and informed of any required corrective action in the next Office action. The objection to the drawings will not be held in abeyance.

***Claim Rejections - 35 USC § 112***

4. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

5. Claims **3, and 4,** are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

6. Claims **3, and 4,** are recites the limitation *"structured document"*, wherein independent claim 1, recites the limitation *"structured sample document"*. There is insufficient antecedent basis for this limitation in the claim.

***Claim Rejections - 35 USC § 103***

7. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

**8. Claims 1 – 8, 11 - 20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cowan et al. ("Rita-An Editor And User Interface For Manipulating Structured Documents"; Pub Date: 1991), in view of Lindbald et al. (Pub. No.: US 2004/0103091 A1; Filing Date: Jun. 13, 2003), and further in view of Brockway et al. (Pub. No.: US 2004/0249795 A1; Filing Date: Jun. 5, 2003).**

**In regards to independent claim 1,** Cowan et al. discloses Rita, an editor and user interface for manipulating *structured documents*, is used for creating and editing *tagged* documents which are structurally correct as they are entered into the computer (Introduction, 2nd paragraph, lines 2-4). The user supplies the *text* of the document and chooses *tags* from a menu, which is constrained by the document grammar so that only structurally correct documents are entered (System Overview, 2nd paragraph, lines 5-7). Cowan et al. further disclose, documents can be adjudged simple or complex on a

number of criteria including: the set of tags that are used, possible attributes or modifiers for the tags, and the use of low-level formatting commands. Simple document usually use a subset of the available tags for document type, and do not use attributes or formatting commands (Editing Complex Documents, 2nd para.; lines 1-4)

Cowan et al. does not disclose expressly, *providing sample data in the form of structured sample documents.*

However, Lindbald et al. teaches, comparing much research and many systems are known for characterizing test document (*structured sample documents*) and analyzing the resulting qualities. For example, a search engine characterizes text documents by their word content, relative placement and occurrence of words and possibly their storage location, and an analysis process selects text document by whether or not they meet some specified search query constraint (section 0056, lines 2-7).

Lindbald et al. does not disclose expressly, *for a selected content element, suggesting an optimal tag according to a tag suggestion procedure.*

However, Brockway et al. discloses a "runHeuristic method" that calls a "compare method" that scans through a tree structure of a document (*structured sample documents*), and through the tree structure of a model document structure, comparing the two trees, incrementing an integer score for all points of positive comparison, and returning a score value (section 0084). The "runHeuristic method" then compares the score with its comparison threshold *(analyzing patterns in the sample data to derive a set of tags suggestions)*, and if the score exceeds the threshold *(evaluating the set of*

*candidate tags according to tag suggestions criteria to determine an optimal tag)*, the "runHeuristic method" concludes that the semantics of its document structure template are good for the current document (section 0086).

Therefore, at the time of the invention, it would have been obvious for a person of ordinary skill in the art to combine Brockway et al. "runHeuristic method" with Cowan et al. user interface for manipulating structured documents, in view of Lindbald et al. teachings, for the benefit providing an optimal tag for a content element within a structured document (Brockway et al., section 0086, lines 3-4).

**In regards to independent claim 11,** Cowan et al. discloses Rita, an editor and user interface for manipulating *structured documents*, is used for creating and editing *tagged* documents which are structurally correct as they are entered into the computer (Introduction, 2nd paragraph, lines 2-4). The user supplies the text of the document and chooses *tags* from a menu, which is constrained by the document grammar so that only structurally correct documents are entered (System Overview, 2nd paragraph, lines 5-7). Cowan et al. also disclose, documents can be adjudged simple or complex on a number of criteria including: the set of tags that are used, possible attributes or modifiers for the tags, and the use of low-level formatting commands. Simple document usually use a subset of the available tags for document type, and do not use attributes or formatting commands (Editing Complex Documents, 2nd para.; lines 1-4). Further, Cowan et al. discloses, that with Rita, structures can be marked with the cursor, then the structures and their accompanying text can be cut and pasted, or written to a file and merged into

another document fragment *(content fragment)* (Structure editing, page 131, last sentence on page).

Cowan et al. does not disclose expressly, *providing a sample structured sample documents.*

However, Lindbald et al. teaches how a collection of XML documents might be decomposed into a forest of subtrees, where each subtree describes a *fragment* within one of the XML documents (section 0040). Lindbald et al. also teaches, comparing much research and many systems are known for characterizing test document (*structured sample documents*) and analyzing the resulting qualities. For example, a search engine characterizes text documents by their word content, relative placement and occurrence of words and possibly their storage location, and an analysis process selects text document by whether or not they meet some specified search query constraint (section 0056, lines 2-7).

Lindbald et al. does not disclose expressly, *for a selected tag, suggesting an optimal content fragment according to a content suggestion procedure.*

However, Brockway et al. discloses a "runHeuristic method" that calls a "compare method" that scans through a tree structure of a document (*structured sample documents*), and through the tree structure of a model document structure, comparing the two trees, incrementing an integer score for all points of positive comparison, and returning a score value (section 0084). The "runHeuristic method" then compares the score with its comparison threshold, and if the score exceeds the threshold *(evaluating the set of content fragment according to content fragment suggestions criteria to*

*determine an optimal content fragment)*, the "runHeuristic method" concludes that the semantics of its document structure template are good for the current document (section 0086).

Therefore, at the time of the invention, it would have been obvious for a person of ordinary skill in the art to combine Brockway et al. "runHeuristic method" with Cowan et al. user interface for manipulating structured documents, in view of Lindbald et al. teachings, for the benefit providing an optimal content fragment for a selected tag, within a structured document (Brockway et al., section 0086, lines 3-4).

**In regards to dependent claims 2, and 7,** Cowan et al. does not disclose expressly, *the tag suggestion criteria comprises satisfying a similarity function.*

However, Lindbald et al. discloses another analysis that can be done on data is a comparison for similarity. Many techniques have been developed to measure similarity among data set, where the data being tested comprises text document, and well-developed techniques that are typically used to determine a similarity measure between two text documents (section 0019, lines 9-11).

Therefore it would have been obvious to a person of ordinary skill in the art to develop tag suggestion criteria, which comprise satisfying a similarity function. The motivation in doing so would have been that most similarity measuring techniques are suitable to properly analyze certain data, such as structure text *(tags)* as might be found in an XML document (section 0020).

**In regards to dependent claim 3,** Cowan et al. discloses, when the user wishes to enter a new document, there are times when a tag set is not adequate, and low-level formatting commands must be introduced in the document. A document grammar can contain a "process specific" tag, which allows the insertion of such formatting commands (page 133, lines 3-5).

**In regards to dependent claim 4,** Cowan et al. discloses, when the user wishes to enter a new document, the cursor is placed on a tag in the structure window, and a menu appears at the bottom of the screen (page 131; 3rd paragraph).

**In regards to dependent claim 5,** Cowan et al. does not disclose expressly, *the structured sample document comprises an XML document having a DTD associated with it.*

However, Lindbald et al. teaches, comparing much research and many systems are known for characterizing test document and analyzing the resulting qualities. Lindbald et al. also teaches, in many cases data set (*structured sample documents*) are XML documents (section 0056, lines 8-9).

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to use a sample structure document in XML format. The motivation in doing so would have been because XML documents are typically defined from Document Type Definitions (DTD).

**In regards to dependent claim 6,** Cowan et al. does not disclose expressly, *the set of tag suggestion includes tree patterns of tags.*

However, Lindbald et al. discloses XML documents are describe as comprising a *tree of tags* elements with some *tag* elements having attributes, and each *tag* element having content (section 0057, lines 10-17). Lindbald et al. illustrates a schematic representation of an XML document, wherein the schematic representation is shown as a *tree* with each node representing an element of the XML document, or an element's content, attribute (section 0042, Fig. 4A.). Each "tag" node is a parent node to a data node (section 0043, lines 14-16; Fig. 4A.).

Therefore, at the time of the invention, it would have been obvious to a person of ordinary skill in the art to combine Lindbald et al. with Cowan et al. for the benefit of nodes *(tag)*, within the tree, having the property that one node is the root node, and all of the other nodes *(tag)* of the set can be reached by following the edges in the orientation direction from the root node (Lindbald et al., section 0044, lines 1-5), to obtain the invention as specified in the claim(s).

**In regards to dependent claim 8,** Cowan et al. does not disclose expressly, *tag suggestion criteria comprises balancing size of tree patterns of tags.*

However, Lindbald et al. discloses XML documents are describe as comprising a *tree of tags* elements with some *tag* elements having attributes, and each *tag* element having content (section 0057, lines 10-17). Lindbald et al. illustrates a schematic representation of an XML document, wherein the schematic representation is shown as

a *tree* with each node representing an element of the XML document, or an element's content, attribute (section 0042, Fig. 4A.). Each "tag" node is a parent node to a data node (section 0043, lines 14-16; Fig. 4A.). Lindbald et al. further disclose decomposition rules need not be specific to tag names, but can specify breaks upon occurrence of other conditions, such as reaching a certain *size* of a subtree or subtree content (*tree pattern of tags)* (section 0049, lines 15-17).

Therefore, at the time of the invention, it would have been obvious to a person of ordinary skill in the art to combine Lindbald et al. with Cowan et al. for the benefit of counting the occurrence of tree pattern of tags within the sample data, to obtain the invention as specified in the claim(s).

**In regards to dependent claim 12, 13, and 15,** Cowan et al., in view of Lindbald et al., does not disclose expressly, *assigning a score to each content fragment in the set of content fragment, wherein the score is a ratio of the number of occurrences of the content fragment.*

However, Brockway et al. discloses a "runHeuristic method" that calls a "compare method" that scans through a tree structure of a document (*structured sample documents*), and through the tree structure of a model document structure, comparing the two trees, incrementing an integer *score* for all points of positive comparison, and returning a *score* value (section 0084). The "runHeuristic method" then compares the *score* with its comparison threshold, and if the score exceeds the threshold *(evaluating the set of content fragment according to content fragment suggestions criteria to*

*determine an optimal content fragment)*, the "runHeuristic method" concludes that the semantics of its document structure template are good for the current document (section 0086). Brockway et al. also discloses an exemplary measure of relevance is a count of how many times a keyword *(content fragment)* occurs in a partition of a structured document (section 0109, lines 11-13).

Therefore, at the time of the invention, it would have been obvious for a person of ordinary skill in the art to combine Brockway et al. "runHeuristic method" with Cowan et al. user interface for manipulating structured documents, in view of Lindbald et al. teachings, for the benefit providing an optimal content fragment for a selected tag, within a structured document (Brockway et al., section 0086, lines 3-4), to obtain the invention as specified in the claim(s).

**In regards to dependent claims 14, and 19,** Cowan et al. does not disclose expressly, *assigning a context to each content fragment in a set of content fragments, wherein context comprises the structural context of the tag surrounding the context fragment.*

However, Lindbald et al. discloses XML documents are describe as comprising a *tree of tags* elements *(structure context of tags)* with some *tag* elements having attributes, and each *tag* element having content *(content fragment)* (section 0057, lines 10-17). Lindbald et al. also illustrates a schematic representation of an XML document, wherein the schematic representation is shown as a *tree* with each node representing an element of the XML document, or an element's content, attribute (section 0042, Fig.

4A.). Each "tag" node is a parent node to a data node (section 0043, lines 14-16; Fig. 4A.).

Therefore, at the time of the invention, it would have been obvious to a person of ordinary skill in the art to combine Lindbald et al. with Cowan et al. for the benefit of assigning a tag to each fragment of the XML document, to obtain the invention as specified in the claim(s).

**In regards to dependent claim 16,** Cowan et al. does not disclose expressly *each content fragment is referenced by a partial path from the structured document root.*

However, Lindbald et al. discloses XML documents are describe as comprising a *tree of tags* elements with some *tag* elements having attributes, and each *tag* element having content (section 0057, lines 10-17). Lindbald et al. illustrates a schematic representation of an XML document, wherein the schematic representation is shown as a *tree* with each node representing an element of the XML document, or an element's content, attribute (section 0042, Fig. 4A.). Each "tag" node is a parent node to a data node (section 0043, lines 14-16; Fig. 4A.).

Therefore, at the time of the invention, it would have been obvious to a person of ordinary skill in the art to combine Lindbald et al. with Cowan et al. for the benefit of nodes *(content fragment)*, within the tree, having the property that one node *(content fragment)* is the root node, and all of the other nodes *(content fragment)* of the set can be reached by following the edges in the orientation direction from the root node

(Lindbald et al., section 0044, lines 1-5), to obtain the invention as specified in the claim(s).

**In regards to dependent claims 17, and 18,** Cowan et al. discloses tags attached to "small items", such as single words, or phases are hidden in terse mode, so as not to disrupt the flow of text *(small linguistic unit)* (page 132, last sentence on page).

Cowan et al. does not disclose expressly, *assigning a score to the small linguistic unit, wherein the score is a ratio of the number of occurrences of the small linguistic unit under the selected tag.*

However, Brockway et al. discloses a "runHeuristic method" that calls a "compare method" that scans through a tree structure of a document (*structured sample documents*), and through the tree structure of a model document structure, comparing the two trees, incrementing an integer *score* for all points of positive comparison, and returning a *score* value (section 0084). The "runHeuristic method" then compares the *score* with its comparison threshold, and if the score exceeds the threshold, the "runHeuristic method" concludes that the semantics of its document structure template are good for the current document (section 0086). Brockway et al. also discloses an exemplary measure of relevance is a count of how many times a keyword *(linguistic unit)* occurs in a partition of a structured document (section 0109, lines 11-13).

Therefore, at the time of the invention, it would have been obvious for a person of ordinary skill in the art to combine Brockway et al. "runHeuristic method" with Cowan et al. user interface for manipulating structured documents, in view of Lindbald et al.

teachings, for the benefit of assigning a score to the small linguistic unit, to obtain the invention as specified in the claim(s).

**In regards to dependent claim 20,** Cowan et al. discloses Rita, an editor and user interface for manipulating *structured documents*, is used for creating and editing *tagged* documents which are structurally correct as they are entered into the computer (Introduction, 2nd paragraph, lines 2-4). The user supplies the *text* of the document and chooses *tags* from a menu, which is constrained by the document grammar so that only structurally correct documents are entered (System Overview, 2nd paragraph, lines 5-7).

9. **Claims 9, and 10 are rejected under 35 U.S.C. 103(a) as being unpatentable over Cowan et al. ("Rita-An Editor And User Interface For Manipulating Structured Documents"; Pub Date: 1991), in view of Lindbald et al. (Pub. No.: US 2004/0103091 A1; Filing Date: Jun. 13, 2003), in view of Brockway et al. (Pub. No.: US 2004/0249795 A1; Filing Date: Jun. 5, 2003), and further in view of Kotsakis ("Structured Information Retrieval In XML Documents"; Pub Date: 2002).**

**In regards to dependent claims 9, and 10,** Cowan et al., in view of Lindbald et al., and

further in view of Brockway et al., does not disclose expressly *a set of tree patterns of tags "ti E T", and a set of C of candidates is a set of all patterns in T.*

However, Kotsakis discloses a system of retrieving structured information in XML documents. Kotsakis discloses a XML document represented in the form of a summary tree *(T)* (Fig. 2A & 2B). The tree is loaded into an index structure where the content data is separated from the path data. The path data is a hierarchy of *tags*, which records every single path in the collection (page 664, right column). Kotsakis uses an algorithm to insert a summary tree in the index structure by storing the structure part of the summary tree into the path index (tags), and the literal part of the element content into a inverted file (page 665, left column). Kotsakis also discloses an UpdateInvertedFile (I, t, c) method that stores the literal content of the node *t* into the inverted file *"I"*. Content(t)

is the literal content of the tag "t" in the summary tree. "c" is the node (tag) in the path index to which all terms in content(t) will be linked (page 665, right column). Kotsakis further discloses a ranking scheme that is divided into two components. The first one defines the term weight in terms if its distribution and the second one in terms of it's structural position. (Ranking, page 666). The path index contains normalized tags. This feature may facilitate *similarity* search *(similarity function)* by content and structure (Index 2 of Conclusion; page 666).

The current invent is similar to Kotsakis in that Kotsakis teaches the concepts used in the current invention. Kotsakis teaches the concept of separating the content data *(set of C of candidates)* from the path data *(tree patterns of tags)*. Kotsakis also teaches the concept of ranking, which would have been obvious to a person of ordinary skill in the art when testing for similarity between the set of optimal tag candidates and the tree patterns of tags.

At the time of the invention, it would have been obvious to a person of ordinary skill in the art to combine Kotsakis with Cowan et al., in view of Lindbald et al., and further in view of Brockway et al. for the benefit of optimizing the process of suggesting the optimal tag according to a tag suggestion procedure, to obtain the invention as specified in the claims.

***Conclusion***

Any inquiry concerning this communication or earlier communications from the examiner should be directed to James J. Debrow whose telephone number is 571-272-5768. The examiner can normally be reached on 8:00-5:00pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather Herndon can be reached on 571-272-4136. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

James Debrow
Examiner
Art Unit 2176

HEATHER R. HERNDON
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100